CMSC 858S: Randomized Algorithms

Fall 2001

Handout 2: Basics regarding expectations and the probabilistic method

1 Some basic properties of the expectation

Here, we will assume for simplicity that all random variables X_i that we consider take on values only from some finite sets of real numbers, but all these facts are true for arbitrary random variables.

We start by proving the all-important *linearity of expectation*: for *arbitrary* random variables $X_1, X_2, \ldots, X_n$ (but which only take on values from some finite sets of real numbers, for simplicity),

$$\mathbf{E}[X_1 + X_2 + \cdots + X_n] = \mathbf{E}[X_1] + \mathbf{E}[X_2] + \cdots + \mathbf{E}[X_n].$$

To show this, we claim that it suffices to consider the case where $n = 2$, and then applying induction suitably; verify this.

So suppose $n = 2$, and that random variables X_1 and X_2 take on values from the finite sets S_1 and S_2 respectively. We wish to show that $\mathbf{E}[X_1 + X_2] = \mathbf{E}[X_1] + \mathbf{E}[X_2]$. For $i = 1, 2$, let $p_i(x)$ denote $\Pr[X_i = x]$; also let $q(x, y)$ denote $\Pr[(X_1 = x) \wedge (X_2 = y)]$. Note that we **cannot** assume that $q(x, y) = p_1(x) \cdot p_2(y)$, since we are not given that X_1 and X_2 are independent; however, we will use the facts

$$p_1(x) = \sum_{y \in S_2} q(x, y), \text{ and} \tag{1}$$

$$p_2(y) = \sum_{x \in S_1} q(x, y). \tag{2}$$

Verify that these simple but important identities are true.

We have

$$\begin{aligned}
\mathbf{E}[X_1] + \mathbf{E}[X_2] &= (\sum_{x \in S_1} x p_1(x)) + (\sum_{y \in S_2} y p_2(y)) \\
&= (\sum_{x \in S_1, y \in S_2} x q(x, y)) + (\sum_{x \in S_1, y \in S_2} y q(x, y)) \quad \text{[by (1) and (2)]} \\
&= \sum_{x \in S_1, y \in S_2} (x + y) \cdot q(x, y) \\
&= \mathbf{E}[X_1 + X_2].
\end{aligned}$$

Thus we have a proof of the linearity of expectation. Note also that for any constant a, $\mathbf{E}[a \cdot X] = a \cdot \mathbf{E}[X]$. Thus, the linearity of expectation is equivalent to the more general statement that for any constants $a_1, a_2, \ldots, a_n$,

$$\mathbf{E}[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i \mathbf{E}[X_i].$$

Exercise: Show that if random variables $X_1, X_2, \ldots, X_n$ are *independent*, then $\mathbf{E}[X_1 \cdot X_2 \cdots X_n] = \mathbf{E}[X_1] \cdot \mathbf{E}[X_2] \cdots \mathbf{E}[X_n]$. Start again with the case $n = 2$, and then extend to all n using induction.

2 Some illustrations of the probabilistic method

There are two basic recipes in applying the probabilistic method. First, suppose we wish to show that in some (finite) set X, there exists some object x that satisfies some predicate $\mathcal{P}$: i.e., $\mathcal{P}(x)$ is true for some $x \in X$. To show this, we construct some probability distribution D over the elements of X, and prove that if Y denotes a random sample from D, then $\Pr[\mathcal{P}(Y) \text{ holds}] > 0$. Second, if we wish to show that there is some $x \in X$ such that $f(x) \geq a$, then we again construct a distribution D, and try to show that $\mathbf{E}[f(Y)] \geq a$. (The case where we want to prove that there is some $x \in X$ such that $f(x) \leq a$ is handled analogously.) It is often not obvious what a suitable choice of D is, but practice makes us more skilled in this. Please refer to *The Probabilistic Method, Second Edition* by N. Alon and J. H. Spencer, Wiley, 2000, for an excellent treatment of the probabilistic method.

Before proceeding, we recall that

$$(n/r)^r \leq \binom{n}{r} \leq (ne/r)^r.$$

Example 1: We wish to show that if n is large enough, then there exist n-vertex graphs with no clique of size more than $2\log_2 n$, and no independent set of size more than $2\log_2 n$. Define $t = 2\log_2 n$ for convenience.

Define the random graph model $G(n,p)$ to be the distribution on n-vertex graphs where we take n vertices that have distinct labels, and *independently* put an edge between each pair of distinct vertices with probability p. Generate a random graph G from $G(n,p)$, where we shall choose p a bit later. Using the union bound, we claim that the probability that G has some clique or some independent set of size more than t, is at most

$$\sum_{S \subseteq \{1,2,\ldots,n\}:\ |S|=t} [p^{\binom{t}{2}} + (1-p)^{\binom{t}{2}}];$$

verify this.

Now substitute $p = 1/2$, and use the above-seen upper bound on $\binom{n}{t}$, to show that this probability bound is (much) lesser than 1, if n is large enough. Thus we see the existence of the desired graphs.

Example 2: The above was an example of the first recipe of the probabilistic method. Now we see two examples of the second recipe.

Given an n-vertex, m-edge graph, we claim that there is a cut with at least $m/2$ edges crossing the cut. To prove this, we generate a random cut by letting each vertex *independently* choose one side of the cut or the other, with probability $1/2$ each. For any given edge (u,v), we can check that the probability of (u,v) being in the cut is $(1/2) \times (1/2) + (1/2) \times (1/2) = 1/2$. Now, by setting up an indicator random variable for this event, and appying the linearity of expectation to the sum, over all edges in the graph, of these indicators, we see that the expected number of edges in the cut is $m/2$. Now apply the second recipe of the probabilistic method.

Next, given a 3-SAT formula–a Boolean disjunctive formula where each clause is a conjunction of exactly 3 literals (with no "degenerate" cases such as "$x_i \wedge x_i$" or "$x_i \wedge \overline{x_i}$" occuring in any given clause)–we show that there is an assignment of Boolean values to the underlying variables such that at least 7/8th of the clauses are satisfied. To do this, independently set each variable to True or False with equal probability. The probability that a given clause is satisfied is seen to be $1 - 1/8 = 7/8$ (why?). Now, by our usual usage of indicator variables and the linearity of expectation, we see that the expected number of satisfied clauses is 7/8th the number of clauses.

An important extension to the basic probabilistic method is the *method of alteration*: construct a random structure, and then alter it a bit if necessary, to satisfy all the required properties. Furthermore, the examples above considered the simple case where every two-way choice gave equal probability (i.e., 1/2) to both choices. This is often by no means optimal: we often set these probabilities to p and $1-p$ for some yet-to-be-determined p, and finally choose the "best" value for p. The following two examples illustrate these two key ideas.

Example 3: Given a graph G with n vertices and $m \geq n/2$ edges, let us show that G has an independent set of size at least $n^2/(4m)$. Choose each vertex independently with probability p; for each edge with both end-points chosen, delete one of these end-points arbitrarily, to yield a final independent set. (Note the initial random choice followed by a deterministic alteration step.) If X denotes the number of initially chosen vertices and Y denotes the number of edges with both end-points chosen, then the expected size of the independent set is at least $\mathbf{E}[X-Y] = \mathbf{E}[X] - \mathbf{E}[Y]$. We employ indicator variables and the linearity of expectation in the usual way to get that $\mathbf{E}[X] = np$ and $\mathbf{E}[Y] = mp^2$. Now, substituting $p = n/(2m)$ to optimize the expression $np - mp^2$, we get the expected size of the final independent set to be at least $n(n/(2m)) - m(n/(2m))^2 = n^2/4m$. (The assumption $m \geq n/2$ was needed to ensure that $n/(2m) \leq 1$.)

Example 4: A dominating set in a graph $G = (V, E)$ is a set $S \subseteq V$ such that each vertex is either in S or has some neighbor in S. Given a graph with minimum degree $\delta \geq 1$, we show that G has a dominating set of size at most $\frac{1+\ln(\delta+1)}{\delta+1} \cdot n$. Like in example 3, choose each vertex independently with probability p for a yet-unspecified p; every vertex such that neither it, nor any of its neighbors got chosen in the above step, is now added to the set of chosen vertices. Note that these chosen vertices now form a dominating set, and verify that the expected cardinality of this dominating set is at most $np + n(1-p)^{\delta+1}$. (The first term is from the initial random choice, and the second term comes from the alteration step.) We now wish to choose p to minimize this. While such a p can be calculated exactly, it is unwieldy; so we first use the fact that $(1-p)^{\delta+1} \leq e^{-p(\delta+1)}$, and set $p = p_0 \doteq \ln(\delta+1)/(\delta+1)$ to minimize $f(p) = np + ne^{-p(\delta+1)}$. So, the expected size of the final dominating set is at most $f(p_0) = \frac{1+\ln(\delta+1)}{\delta+1} \cdot n$, as desired.